



## Bayesian Age-Period-Cohort Modeling and Prediction – BAMP

Volker J. Schmid  
Imperial College London

Leonhard Held  
Universität Zürich

---

### Abstract

The software package **BAMP** provides a method of analyzing incidence or mortality data on the Lexis diagram, using a Bayesian version of an age-period-cohort model. A hierarchical model is assumed with a binomial model in the first-stage. As smoothing priors for the age, period and cohort parameters random walks of first and second order, with and without an additional unstructured component are available. Unstructured heterogeneity can also be included in the model.

In order to evaluate the model fit, posterior deviance, DIC and predictive deviances are computed. By projecting the random walk prior into the future, future death rates can be predicted.

*Keywords:* Bayesian hierarchical models, age-period-cohort models, prediction.

---

## 1. Introduction

In epidemiology, incidence and mortality counts are usually stratified by incidence (or mortality) year and age groups. Figure 1 shows an example dataset for the 5-year-periods 1965–69 to 1995–99 with data for the age groups 30–34 up to 70–74 depicted in a Lexis diagram (Lexis 1875). Often the covariates which cause these incidences cannot be observed directly. A commonly used approach in such a situation is the age-period-cohort (APC) model. Here the incidence rates are analyzed with regard to three different time scales, age (time between birth and death), period (time of death) and cohort (time of birth). This three time scales are surrogate measures, which can give an indication to the underlying causes of a disease Holford (1998): *Age* is often the main factor in such an analysis as it accounts for consistent extrinsic factors. The effect of the *period* accounts for all factors, which affect every person at a time period in history such as pollution or medical advances. The *cohort* effect accounts for events, which affect generations, e.g., malnutrition of children during or after wars or changing habits

like the increasing number of young female smokers (Knorr-Held and Rainer 2001).

In this paper we present the software **BAMP** (**B**ayesian **A**ge-**P**eriod-**C**ohort **M**odeling and **P**rediction), which allows to analyze incidence count data with a Bayesian age-period-cohort model. **BAMP** analyzes incidence or mortality count data with a Bayesian age-period-cohort model and allows to include several features:

- The input data do not have to be on the same time scale, for example period can be in one year intervals and age grouped in five year intervals; however, time scales have to be equidistant,
- **BAMP** allows for prediction of the future number of cases,
- **BAMP** allows retrospective predictions for existing data for model checking purposes,
- **BAMP** can analyze different APC models, i.e., with and without additional global and local heterogeneity, with RW priors of first or second order, and AP and AC models (Clayton and Schifflers 1987).

**BAMP** is available for free download at <http://www.volkerschmid.de/bamp.html>.

There are some graphical routines available in order to

- plot estimated age, period and cohort effects (only for RW1 model)
- compare observed with fitted and predicted rates
- graphically assess the "significance" of the unstructured parameters. This helps to identify unstructured variation in the data which cannot be described by the age, period and cohort parameters alone.

Section 2 describes the theory of Bayesian age-period-cohort models and modifications to the basic model -

and to add additional heterogeneity parameters. An important feature of Bayesian APC models is the possibility of projecting future incidence rates; this is also described in Section 2. Section 3 describes how **BAMP** is used and how different models can be compared. Section 4 describes additional scripts for visualizing results.

## 2. Bayesian APC models

In the following let  $i = 1, \dots, I$  denote the index of the age group,  $j = 1, \dots, J$  the index of the period and  $k = 1, \dots, K$  the index of the birth cohort. The cohort index can explicitly be computed from age group and period index of an incidence, see also Figure 1; if age and period are on the same time scale, the cohort index is

$$k = k(i, j) = (I - i) + j. \quad (1)$$

If age and period are measured on different scales Equation 1 has to be changed accordingly (Holford 1983); as example, if data are give per year, but the age groups cover five years, the cohort index is

$$k = k(i, j) = 5 \cdot (I - i) + j. \quad (2)$$

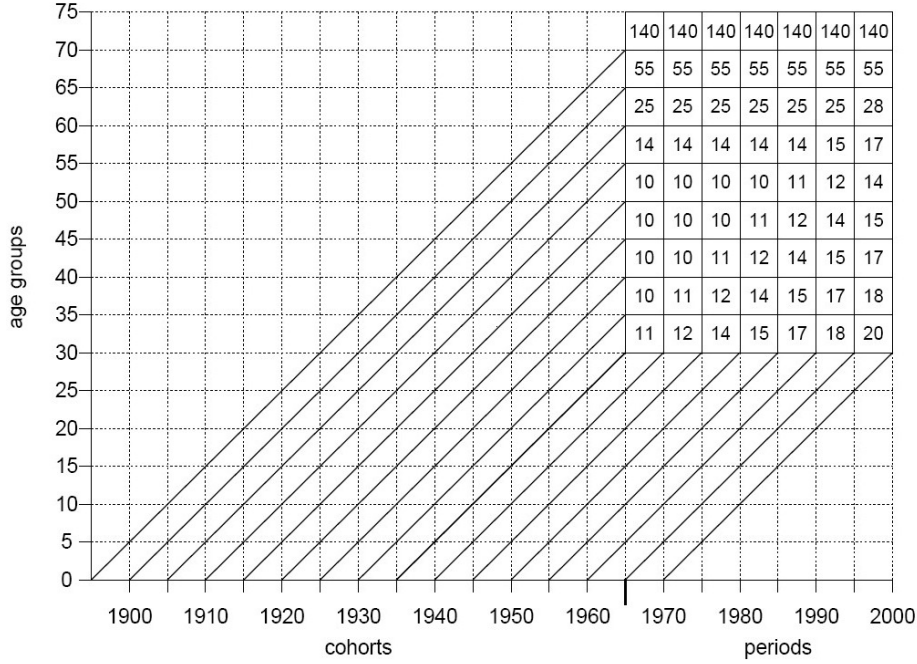


Figure 1: Lexis diagram with an example data set. The Lexis diagram shows number of incidence cases for several years (periods) in different age groups. In the Lexis diagram incidences can be related to the relevant birth cohort. As [Knorr-Held and Rainer \(2001\)](#) point out, cohort groups are overlapping.

In classical APC literature (for an overview see [Heuer 1994](#)) the APC model is often regarded as a log-linear Poisson model. As an alternative a binomial logit model can be formulated (both models are approximately identical): The counts of incidences  $y_{ij}$  in age group  $i$  in period  $j$  follow a binomial distribution with parameters  $p_{ij}$  and  $n_{ij}$ . Here  $n_{ij}$  is the known population size of age group  $i$  at period  $j$  and  $p_{ij}$  is the unknown incidence probability. The logit of the incidence probability is decomposed in an intercept  $\mu$ , age effect  $\theta_i$ , period effect  $\phi_j$  and cohort effect  $\psi_k$ ,

$$y_{ij} \sim B(n_{ij}, p_{ij}), \quad (3)$$

$$\eta_{ij} = \log \left( \frac{p_{ij}}{1 - p_{ij}} \right) = \mu + \theta_i + \phi_j + \psi_k. \quad (4)$$

## 2.1. Non-identifiability

To make the intercept  $\mu$  identifiable, the following restrictions have to be imposed

$$\sum_i \theta_i = \sum_j \phi_j = \sum_k \psi_k = 0. \quad (5)$$

However, the APC parameters in this model are still not identifiable. It can be seen from (1) that for every transformation

$$\mu \rightarrow \mu - c \cdot I; \quad \theta_i \rightarrow \theta_i + c \cdot i; \quad \phi_j \rightarrow \phi_j - c \cdot j; \quad \psi_k \rightarrow \psi_k + c \cdot k \quad (6)$$

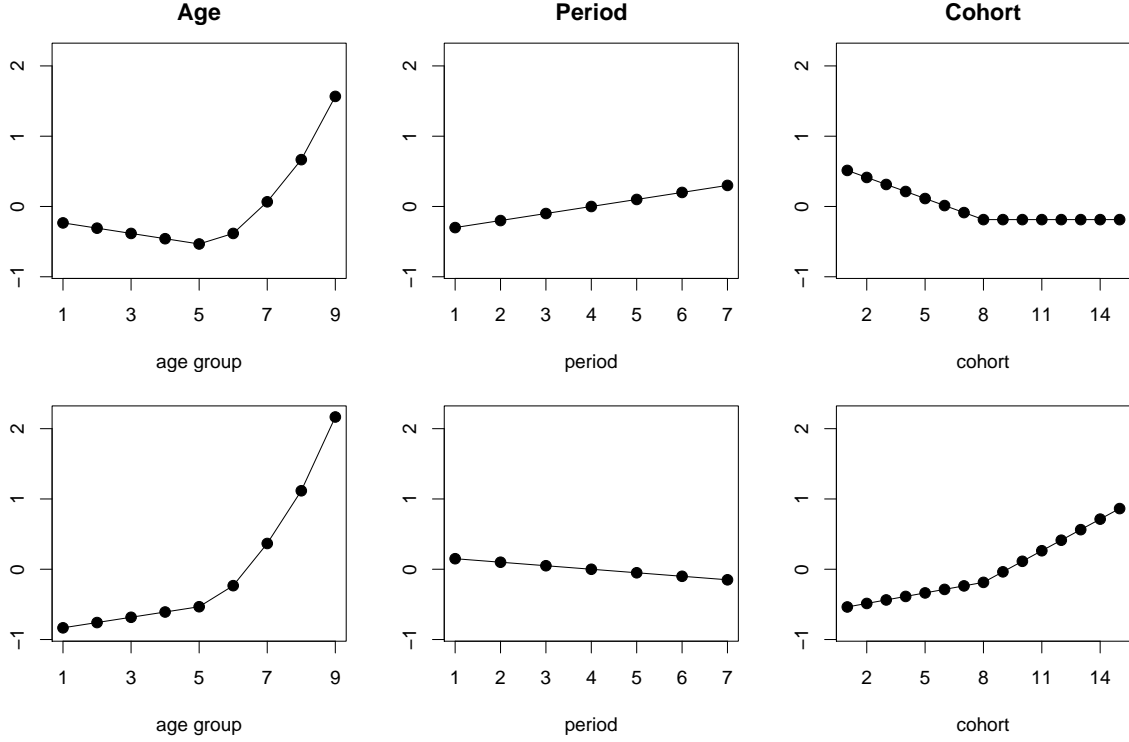


Figure 2: Two transformations of the estimated effects for the example dataset in Figure 1; top row:  $\mu = -6.1$ , bottom row:  $\mu = -5.2$ .

for all  $i, j$  and  $k$  with any  $c \in \mathbb{R}$  the linear predictor  $\eta_{ij}$  is not changed. Therefore the linear predictor can always be identified and interpreted, but age, period and cohort effects cannot. Figure 2 shows the problem of non-identifiability. The figure shows two possible parameter sets for the example dataset in Figure 1. Depending on the transformation, the period effect is decreasing and the cohort effect is increasing or *vice versa*. However, the change in trend at cohort  $k = 8$  remains unchanged. This is, because the non-identifiability only affects linear trends; change points and other non-linear trends can be identified (up to the linear trend) – like the non-linear effect of age in the example. Therefore the interpretation of the estimated APC effects has to be restricted to non-linear trends. **BAMP** provides second differences for this purpose.

## 2.2. Bayesian hierarchical model

**BAMP** uses a Bayesian hierarchical approach for the APC model. Following [Berzuini and Clayton \(1994\)](#) random walk (RW) priors of different orders are used for the APC parameters  $\theta$ ,  $\phi$  and  $\psi$ . The RW1 prior assumes a constant trend over the time scale, whereas the RW2 prior assumes a linear time trend. Therefore the order of the random walk should be chosen depending on whether a constant or a linear time trend can be assumed. However, in the APC model the RW prior also acts as a stochastic restriction: With a RW1 the first-order differences of an effects are stochastically restricted to zero; this makes the APC effects

identifiable. With a RW 2 the second-order differences are stochastically restricted to zero; this restriction however is too weak to assure identifiability.

In the following we describe random walk priors for the period effect  $\phi$  – the age and the cohort effects  $\theta$  and  $\psi$  can be treated analogously. For random walks of first order (RW 1) we assume

$$\begin{aligned} p(\phi_1) &\propto \text{const.} \\ \phi_j | \phi_{j-1} &\sim N(\phi_{j-1}, \kappa^{-1}), \text{ for } j = 2, \dots, J \end{aligned} \quad (7)$$

with  $\kappa^{-1}$  a precision parameter; for random walks of second order (RW 2) we assume

$$\begin{aligned} p(\phi_1) = p(\phi_2) &\propto \text{const.} \\ \phi_j | \phi_{j-1}, \phi_{j-2} &\sim N(2\phi_{j-1} - \phi_{j-2}, \kappa^{-1}) \text{ for } j = 3, \dots, J. \end{aligned} \quad (8)$$

The joint distribution of  $\Phi = (\phi_1, \dots, \phi_J)$  is

$$p(\Phi) \propto \lambda^{rg(\mathbf{R})/2} \exp\left(-\frac{\lambda}{2} \Phi' \mathbf{R} \Phi\right),$$

with  $\mathbf{R}$  the precision matrix of RW 1 (Clayton 1996) or RW 2 (Berzuini, Clayton, and Bernardinelli 1993).

The precision parameter  $\kappa$  is a smoothing parameter and will be estimated simultaneously in the model. The higher the precision, the smoother the estimated parameter vector. A Gamma distribution  $\text{Ga}(a, b)$  is used for the precision parameters. Often  $a = 1$  and  $b = 0.001$  or  $a = b = 0.001$  is chosen. When using a RW 2 prior it is often advisable to use a smaller value for  $b$  in order to get smooth estimates (e.g.,  $b = 10^{-5}$ ).

For the intercept  $\mu$  **BAMP** uses a flat prior

$$p(\mu) \propto \text{const.}$$

Identifiability of the intercept is guaranteed with the restrictions stated in (5).

In the Bayesian setting a random walk of first order imposes a stochastic constraint: the RW 1 model will prefer the one transformation Equation 6, where the quadratic first differences are minimal – i.e., the APC effects are kept as constant as possible. Therefore the APC effects can be identified by the software with this model, nevertheless the restriction in interpretation of the resulting estimations still apply.

### 2.3. Projection

Whereas RW 1 implies a constant time trend, RW 2 implies a linear time trend. In both cases the priors can be used to gain projections for future time points. The projected effects can be computed by

$$\begin{aligned} \phi_{J+1} &\sim N(\phi_J, \kappa^{-1}) && \text{for RW 1,} \\ \phi_{J+1} &\sim N(2\phi_J - \phi_{J-1}, \kappa^{-1}) && \text{for RW 2.} \end{aligned} \quad (9)$$

The Bayesian APC model allows to extrapolate period and cohort effects. The age effect can also be extrapolated, however this is usually not necessary, as data cover all age groups of interest.

In order to get the projection of a future rate  $p_{i,J+1}$  the projected period and cohort effects from Equation 9 are applied to Equation 4:

$$\eta_{i,J+1} = \mu + \theta_i + \phi_{J+1} + \psi_{k(i,J+1)}$$

and

$$p_{i,J+1} = \frac{\exp(\eta_{i,J+1})}{1 + \exp(\eta_{i,J+1})}.$$

The projection of future cases can then be computed using the binomial distribution Equation 3; however the population size  $n_{i,J+1}$  must be known or projected by an adequate method to estimate the number of cases. Further future rates  $p_{i,J+2}, p_{i,J+3}, \dots$  can be derived analogously.

## 2.4. Inference

The full conditional of the APC parameter vectors are non-standard distributions. Therefore **BAMP** uses a Metropolis-Hastings (MH) algorithm to sample from the posterior. Useful Gaussian proposals for the MH algorithm can be derived using Taylor approximation of second degree (Gamerman 1997; Schmid 2004). **BAMP** uses an algorithm proposed by Rue (2001) for effective sampling from multivariate normal distributions.

The full conditional of the hyper parameter  $\kappa$  is a Gamma distribution with parameters  $a + rg(\mathbf{R})/2$  and  $b + \frac{1}{2}\mathbf{\Phi}'\mathbf{R}\mathbf{\Phi}$ . Also, the full conditionals of  $\tau$  and  $\nu$  are Gamma distributions; therefore the precisions can be sampled using Gibbs steps.

## 2.5. Additional global heterogeneity

Knorr-Held and Rainer (2001) propose an additional parameter  $z_{ij}$  for the APC model. This parameter accounts for heterogeneity which can not be explained by the APC effects, like overdispersion. The model has the following form:

$$\eta_{ij} = \mu + \theta_i + \phi_j + \psi_k + z_{ij}. \quad (10)$$

A Gaussian distribution

$$z_{ij} \sim N(0, \delta^{-1})$$

is used as prior for the additional heterogeneity with a  $\text{Ga}(a_\delta, b_\delta)$  prior for the precision parameter  $\delta$ .

This model has a interesting technical advantage. Using a reparametrization as proposed by Besag, Green, Higdon, and Mengersen (1995), the full conditionals of the APC effects are Gaussian distributions, such that Gibbs step can be applied (Knorr-Held and Rainer 2001).

## 2.6. Additional heterogeneity on a time scale

In some analyzes using APC models, the estimated effects are too rough because of additional heterogeneity in the effects or outliers in the data. Therefore not only the variance of these effects is overestimated, but also the credibility intervals of the projection will be too broad.

A solution for this can be to introduce additional parameters for heterogeneity on one or more time scales. Even in the model with additional global heterogeneity parameter  $z_{ij}$  (Section 2.5) an additional heterogeneity parameter may be useful, in order to keep the variance of  $z_{ij}$  small.

Let us assume that the period effect a priori follows

$$\phi_j \sim N(\phi_j^*, \lambda_1) \text{ for } j = 1, \dots, J,$$

with a random walk prior with precision  $\lambda_2$  for  $\Phi^* = (\phi_1^*, \dots, \phi_J^*)'$ . Let  $\tilde{\Phi} = (\tilde{\phi}_1, \dots, \tilde{\phi}_{(2J)})'$  be the joint vector of  $\Phi$  and  $\Phi^*$  with

$$\begin{aligned} \tilde{\phi}_{(2j)-1} &= \phi_j \\ \tilde{\phi}_{(2j)} &= \phi_j^* \end{aligned}$$

for  $j = 1, \dots, J$ . Then the joint prior distribution of  $\phi$  and  $\phi^*$  can be formulated as prior distribution of  $\tilde{\phi}$ :

$$p(\tilde{\phi}) \propto \exp \left[ -\frac{1}{2} \tilde{\Phi}' (\lambda_1 \mathbf{I} \otimes \mathbf{A} + \lambda_2 \mathbf{R} \otimes \mathbf{B}) \tilde{\Phi} \right]$$

with

$$\mathbf{A} = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \text{ and } \mathbf{B} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

and  $\otimes$  the Kronecker matrix product.

An analogous modification of the prior can be formulated for cohort and age, however, in practice there is rarely a need for additional heterogeneity for age. Models with additional heterogeneity on more than one time scale are possible, but in these models the stochastic restriction, which comes with the use of RW 1 priors, does not longer apply.

### 3. Using BAMP

**BAMP** is a stand alone software. All necessary settings are loaded from a `.ini` file, by default called `bamp.ini`. Before executing **BAMP**, edit the `bamp.ini` file or write your own `.ini` file (see Section 3.1.1). Execute the software by typing

```
bamp your-ini-file.ini
```

in the command line. On window-based operating systems, **BAMP** can be executed with a double click on the **BAMP** icon. **BAMP** will then use the `bamp.ini` file by default. On Microsoft Windows systems the file `cygwin1.dll` has to be in the same folder as `bamp.exe`.

#### 3.1. Input files

##### *.ini file*

All parameters and input files are specified in a `.ini` file, by default called `bamp.ini`. Tables 1 to 2 show the variables which can be defined in the `.ini` file. The last column of the tables show the default values. Lines signed with † are mandatory, lines with ‡ are mandatory if applicable (eg. parameter for the gamma priors must be given, unless the corresponding effect is discarded from the model). Each row in the `*.ini` file consist of the name of the variable and the value, separated by a “:”. The order of the lines is arbitrary.

<b>cases file</b>	path and name of cases data file	†
<b>population file</b>	path and name of population data file	†
<b>dataorder</b>	order of data in input files (see Paragraph <i>Data files</i> )	0
<b>output folder</b>	path to folder for output files	†
<b>temp folder</b>	path to folder for temporary files and samples	†
<b>output file</b>	file name for general software output (0: console)	0
<b>number of agegroups</b>	number of age groups	†
<b>number of periods</b>	number of periods	†
<b>periods per agegroup</b>	number of period units per age group (e.g., years per age group if periods are per year)	†
<b>number of predictions</b>	number of periods for which rates should be projected	0
<b>predictions</b>	0: no projection at all 1: prediction of rates 2: prediction of cases for comparison with given cases 3: prediction of cases with given population	0
<b>number of iterations</b>	number of iterations of MCMC algorithm (including burn-In)	†
<b>burn in</b>	number of iterations for burn-in (including tuning)	†
<b>step</b>	thinning parameter	†
<b>tuning</b>	number of iterations for tuning	†
<b>samples</b>	1 samples saved for further analysis, 0 else	0
<b>lock</b>	0: overwrite existing samples, 1 else	1
<b>deviance</b>	0: do not compute deviance 1: compute deviance 2: compute deviance also for projected cases (only if predictions: 2)	0

Table 1: Variables in the .ini file – part 1.

### Data files

Two files containing data are necessary: the **cases file** contains the number of disease cases, the **population file** the number of persons under risk. Both files must be in the form of a matrix with  $T$  rows and  $J$  columns. If **dataorder:1**, the data files will be transposed (i.e. the files are matrices with  $J$  rows and  $T$  columns). To predict rates for existing data (**predictions:2**) for  $S$  periods both files must have  $T + S$  rows ( $T + S$  columns for **dataorder:1**). For given population data (**predictions:3**), the **population file** must have  $T + S$  rows ( $T + S$  columns for **dataorder:1**). In all cases redundant rows will be ignored.

### Priors for APC effects

For the age, period and cohort effects the following priors are available, set **age block**, **period block** and **cohort block** accordingly.

0: Discard effect from the model (for AP and AC model)



age block	prior of age effect, see Paragraph <i>Priors for APC effects</i>	0
age hyperpar. a	parameters $a$ and $b$ of the gamma prior of the precision	†
age hyperpar. b	of the age effect	†
age hyperpar.2 a	parameters $a$ and $b$ of the gamma prior of the precision	†
age hyperpar.2 b	of the unstructured age effect	†
period block	prior of period effect, see Paragraph <i>Priors for APC effects</i>	0
period hyperpar. a	parameters $a$ and $b$ of the gamma prior of the precision	†
period hyperpar. b	of the period effect	†
period hyperpar.2 a	parameters $a$ and $b$ of the gamma prior of the precision	†
period hyperpar.2 b	of the unstructured period effect	†
cohort block	prior of the cohort effect, see Paragraph <i>Priors for APC effects</i>	0
cohort hyperpar. a	parameters $a$ and $b$ of the gamma prior of the precision	†
cohort hyperpar. b	of the cohort effect	†
cohort hyperpar.2 a	parameters $a$ and $b$ of the gamma prior of the precision	†
cohort hyperpar.2 b	of the unstructured cohort effect	†
z mode	0: with, 1: without pixel-wise heterogeneity parameter	1
z hyperpar. a	parameters $a$ and $b$ of the gamma prior of the precision of the	†
z hyperpar. b	pixel-wise heterogeneity parameter	†
quantile 1	quantiles of the posterior for output (-1: none)	-1
quantile 2		-1
quantile 3		-1
quantile 4		-1
quantile 5		-1
period covariate	path and name of the covariable data file for period	†
period start	first time point to use from covariables data set	†
cohort covariate	path and name of the covariable data file for cohort	†
cohort start	first time point to use from covariable data set	†

Table 2: Variables in the .ini file – part 2.

- 1: RW 1 prior
- 2: RW 2 prior
- 3: RW 1 prior with additional unstructured heterogeneity (see Section 2.6)
- 4: RW 2 prior with additional unstructured heterogeneity (see Section 2.6)
- 8: RW 1 prior with use of covariates (see Section 3.4)
- 9: RW 2 prior with use of covariates (see Section 3.4)

### 3.2. Output files

The MCMC algorithm produces samples from the posterior distribution of all unknown parameters. These samples are written to the **temp** folder. After finishing the MCMC run, the software calculates quantiles of the samples from the posterior distribution. These quantiles

will be written in the `output` folder. The variables `quantile 1` to `quantile 5` specify the quantiles, which will be calculated. Variables set to `-1` will be omitted.

The output consists of the following files:

- `theta.txt`, `phi.txt`, `psi.txt`: Quantiles of the age, period, cohort effect; each quantile is a row (these will only be produced for RW 1 priors)
- `theta2.txt`, `phi2.txt`, `psi2.txt`: Quantiles of second differences of the age, period, cohort effects (these can always be identified and interpreted)
- `hyper.txt`: Quantiles of the precision parameters of age, period and cohort effects followed by the precision of the unstructured “random effects”, name of parameter and quantiles in each row
- `z.txt`: Quantiles of pixel-wise heterogeneity parameter, each block is one quantile of  $T$  rows with  $J$  columns
- `my.txt`: Quantiles of the intercept parameter  $\mu$
- `devianz.txt` (if `deviance > 0`): Quantiles of the deviance of the model
- `devianz_prog.txt` (if `prediction:2` and `deviance:2`): Quantiles of the deviance for the predicted periods (each row one period)
- `dic.txt`: deviance, effective number of parameters and DIC, see Spiegelhalter, Best, Carlin, and van der Linde (2002)
- `pr.txt`: Quantiles of predicted rates
- `y_pred.txt`, `ysum_pred.txt` (if `prognosis:2,3`, see Section 3.3): Quantiles of predicted cases per age groups and summed up
- `beta_phi.txt`, `beta_psi.txt`: period and cohort effect if covariates are used

### 3.3. Model fit, predictions

In order to evaluate the goodness of the model fit, **BAMP** calculates the posterior deviance. The deviance is defined as:

$$D = 2 \cdot \sum_{j,t} [l(\hat{y}_{jt}) - l(y_{jt})]$$

with  $l(\hat{y}_{jt})$  the individual log-likelihood of the model and  $l(y_{jt})$  the maximum individual log-likelihood achievable. The smaller the deviance, the better the fit of the model. However, the model fit will typically be good for any model as long the unstructured parameters are included. To compare the performance of different models, **BAMP** also provides the deviance information criterion DIC (Spiegelhalter *et al.* 2002) and the predictive deviance, see below.

To make predictions of further death rates, set `predictions:1`. Set `number of predictions` to the number of periods for which you want the projection. If you have population data for these periods, set `predictions:3`, the software will then provide predicted cases. In order to test the model you can also make prediction for existing data. Then the input files must

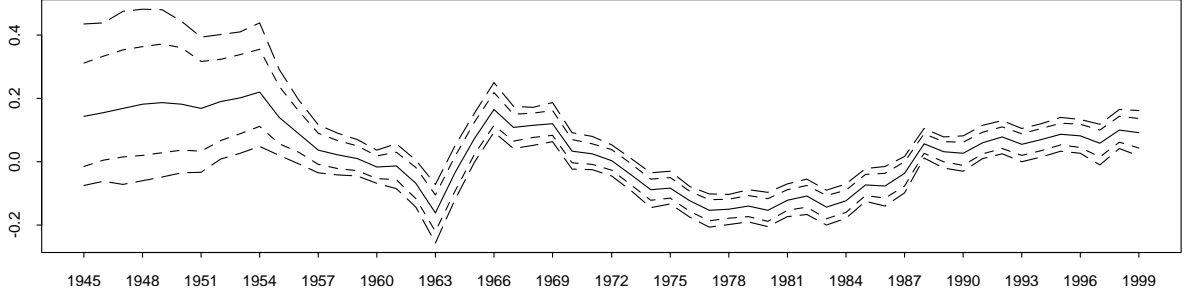


Figure 3: Estimated time variable covariate effect in model (11).

include this data, the matrices have to be of dimension  $J$  times  $(T + S)$  (see Section 3.1). Set `deviance:2` and **BAMP** will calculate the predictive deviance for every projected period. The predictive deviance at time  $t$  is defined as

$$D_t = 2 \cdot \sum_j \left[ l(\hat{y}_{jt}) - l(y_{jt}) \right]$$

with  $\hat{y}_{jt}$  the predicted number of cases in age group  $j$  at period  $t$ . The predictive deviance is a monotonic transformation of the *logarithmic scoring rule* as suggested initially by Good (1952), see also Gneiting and Raftery (2006), to quantify the quality of probabilistic forecasts.

### 3.4. Using covariates

By including a time-variable covariate effect, the Bayesian APC model might be improved:

$$\eta_{ij} = \mu + \theta_i + \beta_j x_{j-L} + \psi_k + z_{ij} \quad (11)$$

with a random walk prior for the regression parameter  $\beta = (\beta_1, \dots, \beta_J)$ . The covariate  $\mathbf{x} = (x_{1-L}, \dots, x_{J-L})$  enters with a time lag  $L$ .

The model is based on the idea, that one of the main effects can be explained by a covariate measure. In an analysis of lung cancer mortality data in West Germany Knorr-Held and Rainer (2001) present a model, where the period effect for females is explained by smoking habits, with a lag of 20–25 years. In a way, the substitution of one main effect with a covariate can be seen as a different type of restriction, where the model is restricted, so that one effect is linear dependent on the covariate.

However, it is not clear, if and how the covariates have to be standardized. If the covariate is near to zero, the credibility intervals get very broad, as Figure 3 shows for a lung disease dataset with the estimated amount of consumed tobacco as covariate, here the values are near to zero in the first years. If the covariate is exactly zero at one time point, the adjoined parameter indeed can not be estimated.

**BAMP** allows to include covariates in the age-period-cohort model. Covariate effects can either replace period or cohort effects. Lets assume we have covariates  $x_t$  for periods  $t = 1 \dots T$ . We can introduce this covariate by replacing the period effect  $\phi$  with  $c_\phi \cdot x_t \beta_t^{(\phi)}$  with  $c_\phi = \frac{T}{\sum x_t}$ . For  $\beta^{(\phi)}$ , a random walk of first or second order can be chosen. In order to include covariates, `period block` has to be set to 8 (RW1) or 9 (RW2). The variable `period`

`covariate` gives the file name of the covariate data. The data have to be separated by space, tabulator or newline. The variable `period start` specifies, which line of the covariate file matches the first period of the cases/population data. Covariates for cohorts can equivalently be included; also both period and cohort effect can be replaced by covariates.

## 4. S-PLUS and R code for graphical output

The graphical routines for S-PLUS ([Insightful Corp. 2003](#)) and R ([R Development Core Team 2007](#))

- plot estimated age, period and cohort effects (for RW1 priors only!)
- compare observed and fitted rates
- plot projected rates
- assess the "significance" of the unstructured parameters

Download the code from the **BAMP** homepage and edit the first lines as appropriate:

- `startingage`: the first age of the first age group of the data
- `startingperiod`: the first period of the data
- `yearsperperiod`: how man years are one period
- `inifile`: name of the `.ini` file

To start, copy the code to the directory where the `.ini` file is located, change to this directory and type

```
Splus < bamp.splus    resp.    R --nosave < bamp.R
```

or start in S-PLUS or R via

```
source(path//bamp.S)    resp.    source(path//bamp.R)
```

The resulting graphic file is named `bamp.ps` in the output directory.

## 5. Summary

Bayesian age-period-cohort models are a useful tool to analyze mortality and incidence rates stratified by age and period. A RW 1 prior for all APC effects implies a stochastic restriction, solving the identifiability problem. However, with RW 2 priors the main effects can not be identified. Whatever RW model is used, the linear trends can not be interpreted without considering the stochastic restrictions implied by the model, however, break points and other non-linear changes can be interpreted. Therefore APC models can be a first step for further analysis of the causes of incidence.

Several recent papers, including the work of exploit the use of Bayesian inference in APC modeling. A Bayesian APC models with RW 2 priors are used in [Bray, Brennan, and Boffetta \(2000\)](#) and [Baker and Bray \(2005\)](#) for projections of cancer mortality, [Knorr-Held and Rainer \(2001\)](#) present a model with overdispersion in a lung cancer study and exploit the use of covariates in a APC model, and [Hansell, Held, Best, Schmid, and Aylin \(2003\)](#) and [Lopez, Shibuya, Rao, Mathers, Hansell, Held, Schmid, and Buist \(2006\)](#) analyze COPD mortality trends with Bayesian APC model. These papers not only show the methodology of Bayesian APC models, but also the interpretation of results and of different possible models.

Predictions in the Bayesian APC model can easily be made by carrying forward the autoregressive priors. The prior has to be chosen carefully: predictions with RW 1 models may have a bias, if the assumption of constant time trends does not apply; models with RW 2 priors tend to have broader credibility intervals, see also the results of [Knorr-Held and Rainer \(2001\)](#) and [Schmid and Held \(2004\)](#).

**BAMP** provides a variety of different APC models. Global heterogeneity can be in- or excluded, additional heterogeneity on time scales is possible; one or even two of the APC effects can be discarded from the model. The different models can be compared via deviance, DIC and predictive qualities. **BAMP** is available for free download at <http://www.volkerschmid.de/bamp.html>.

## References

- Baker A, Bray I (2005). “Bayesian Projections: What Are the Effects of Excluding Data from Younger Age Groups?” *American Journal of Epidemiology*, **162**, 798–805.
- Berzuini C, Clayton D (1994). “Bayesian Analysis of Survival on Multiple Time Scales.” *Statistics in Medicine*, **13**, 823–838.
- Berzuini C, Clayton D, Bernardinelli L (1993). “Bayesian Inference on the Lexis Diagram.” *Bulletin of the International Statistical Institute*, **50**, 149–164.
- Besag JE, Green PJ, Higdon DM, Mengersen KL (1995). “Bayesian Computation and Stochastic Systems.” *Statistical Science*, **10**, 3–66.
- Bray I, Brennan P, Boffetta P (2000). “Projections of Alcohol- and Tobacco-related Cancer Mortality in Central Europe.” *International Journal of Cancer*, **87**, 122–128.
- Clayton D, Schifflers E (1987). “Models for Temporal Variation in Cancer Rates. I: Age-Period and Age-Cohort Models.” *Statistics in Medicine*, **6**, 449–467.
- Clayton DG (1996). “Generalized Linear Mixed Models.” In WR Gilks, S Richardson, DJ Spiegelhalter (eds.), “Markov Chain Monte Carlo in Practice,” pp. 275–301. Chapman & Hall, London.
- Gamerman D (1997). “Sampling from the Posterior Distribution in Generalized Linear Mixed Models.” *Statistics and Computing*, **7**, 57–68.
- Gneiting T, Raftery A (2006). “Strictly Proper Scoring Rules, Prediction, and Estimation.” *Journal of the American Statistical Association*, **107**, 359–378.

- Good I (1952). “Rational Decisions.” *Journal of the Royal Statistical Society B*, **14**, 107–14.
- Hansell A, Held L, Best N, Schmid V, Aylin P (2003). “Did the 1956 Clean Air Act Make a Difference to COPD Mortality Trends 1950–1999?” *Thorax*, **58**(S3), 74–75.
- Heuer C (1994). *Analyse von Zeittrends in Mortalitäts- und Inzidenzraten mit Alters-Perioden-Kohortenmodellen und Regressionssplines*. Ph.D. thesis, Universität Dortmund.
- Holford T (1983). “The Estimation of Age, Period and Cohort Effects for Vital Rates.” *Biometrics*, **39**, 311–324.
- Holford T (1998). “Age-Period-Cohort Analysis.” In P Armitage, T Colton (eds.), “Encyclopedia of Biostatistics,” volume 1, pp. 82–99. John Wiley & Sons, Chichester.
- Insightful Corp (2003). *S-PLUS Version 6.2*. Seattle, WA. URL <http://www.insightful.com/>.
- Knorr-Held L, Rainer E (2001). “Projections of Lung Cancer Mortality in West Germany: A Case Study in Bayesian Prediction.” *Biostatistics*, **2**, 109–129.
- Lexis W (1875). *Einleitung in die Theorie der Bevölkerungsstatistik*. Trübner: Straßburg.
- Lopez A, Shibuya K, Rao C, Mathers C, Hansell A, Held L, Schmid V, Buist S (2006). “Chronic Obstructive Pulmonary Disease: Current Burden and Future Projections.” *European Respiratory Journal*, **27**, 397–412.
- R Development Core Team (2007). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.
- Rue H (2001). “Fast Sampling of Gaussian Markov Random Fields.” *Journal of the Royal Statistical Society B*, **63**, 325–338.
- Schmid V (2004). *Bayesianische Raum-Zeit-Modellierung in der Epidemiologie*. München: Dr.Hut Verlag. Ph.D. thesis, Fakultät für Mathematik, Informatik und Statistik der Ludwig-Maximilians-Universität München.
- Schmid V, Held L (2004). “Bayesian Extrapolation of Space-Time Trends for Cancer Registry Data.” *Biometrics*, **60**, 1034–1042.
- Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A (2002). “Bayesian Measures of Model Complexity and Fit.” *Journal of the Royal Statistical Society B*, **64**, 583–639.

**Affiliation:**

Volker J. Schmid  
Institute of Biomedical Engineering  
Imperial College London  
London SW7 2AZ, United Kingdom  
E-mail: [stats@volkerschmid.de](mailto:stats@volkerschmid.de)  
URL: <http://www.volkerschmid.de/>